



AFRL-RH-WP-TP-2014-0040

Optometric Measurements Predict Performance but not Comfort on a Virtual Object Placement Task with a Stereoscopic 3D Display

John P. McIntire*, Steven T. Wright, Lawrence K. Harrington***, Paul R. Havig*, Scott N. J. Watamaniuk****, and Eric L. Heft***

***USAF AFMC 711 HPW/RHCV, 2255 H St, B248 R300, Wright-Patterson AFB OH 45433-7022**

****USAF AFMC USAFSAM/FECO, 2510 5th St, B840, Wright-Patterson AFB OH 45433-7913**

*****Ball Aerospace & Technology Corp., 2875 Presidential Dr Ste 180, Fairborn OH 45324-6269**

******Wright State U., Psychology Dept, 3640 Col Glenn Hwy, 335 Fawcett Hall, Dayton OH 45435**

September 2014

TECHNICAL PAPER

Distribution A: Approved for public release; distribution unlimited.

STINFO COPY

**AIR FORCE RESEARCH LABORATORY
711 HUMAN PERFORMANCE WING
HUMAN EFFECTIVENESS DIRECTORATE
WRIGHT-PATTERSON AIR FORCE BASE OH 45433
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE**

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

AFRL-RH-WP-TP-2014-0040 HAS BEEN REVIEWED AND IS APPROVED FOR
PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//signed//

DARREL G. HOPPER
Program Manager
Battlespace Visualization Branch

//signed//

JEFFREY L. CRAIG
Chief, Battlespace Visualization Branch
Warfighter Interface Division

//signed//

WILLIAM E. RUSSELL
Chief, Warfighter Interface Division
Human Effectiveness Directorate
711 Human Performance Wing

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
<p>The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</p>					
1. REPORT DATE (DD-MM-YY) 16-09-2014		2. REPORT TYPE Interim (Technical Paper)		3. DATES COVERED (From - To) 1 May 2013 – 12 Feb 2014	
4. TITLE AND SUBTITLE Optometric Measurements Predict Performance but not Comfort on a Virtual Object Placement Task with a Stereoscopic 3D Display				5a. CONTRACT NUMBER FA8650-08-D-6801-0050	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER Multiple	
6. AUTHOR(S) John P. McIntire*, Steven T. Wright**, Lawrence K. Harrington***, Paul R. Havig*, Scott N. J. Watamaniuk****, and Eric L. Heft*				5d. PROJECT NUMBER 5239	
				5e. TASK NUMBER 11	
				5f. WORK UNIT NUMBER H0CK (Historic: 53291102)	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) *USAF AFMC 711 HPW/RHCV, 2255 H St, B248 R300, Wright-Patterson AFB OH 45433-7022 **USAF AFMC USAFSAM/FECO, 2510 5 th St, B840, Wright-Patterson AFB OH 45433-7913 ***Ball Aerospace & Technology Corp., 2875 Presidential Dr Ste 180, Fairborn OH 45324-6269 ****Wright State U., Psychology Dept, 3640 Col Glenn Hwy, 335 Fawcett Hall, Dayton OH 45435				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Material Command Air Force Research Laboratory 711 Human Performance Wing, Human Effectiveness Directorate Warfighter Interface Division, Battlespace Visualization Branch Wright-Patterson Air Force Base OH 45433-7022				10. SPONSORING/MONITORING AGENCY ACRONYM(S) USAF AFMC 711 HPW/RHCV	
				11. SPONSORING/MONITORING AGENCY REPORT NUMBER(S) AFRL-RH-WP-TP-2014-0040	
12. DISTRIBUTION/AVAILABILITY STATEMENT Distribution A: Approved for public release; distribution unlimited. 88ABW Cleared 9/9/2013; 88ABW-2013-3985.					
13. SUPPLEMENTARY NOTES Report contains color.					
14. ABSTRACT Twelve participants were tested on a simple virtual object precision placement task while viewing a stereoscopic 3D (S3D) display. Inclusion criteria included uncorrected or best corrected vision of 20/20 or better in each eye and stereopsis of at least 40 arc sec using the Titmus stereo test. Additionally, binocular function was assessed, including measurements of distant and near phoria (horizontal and vertical) and distant and near horizontal fusion ranges using standard optometric clinical techniques. Before each of six 30 minute experimental sessions, measurements of phoria and fusion ranges were repeated using a Keystone View Telebinocular and an S3D display, respectively. All participants completed experimental sessions in which the task required the precision placement of a virtual object in depth at the same location as a target object. Subjective discomfort was assessed using the Simulator Sickness Questionnaire (SSQ). Individual placement accuracy in S3D trials was significantly correlated with several of the binocular screening outcomes: viewers with larger convergent fusion ranges (measured at near distance), larger total fusion ranges (convergent plus divergent ranges, measured at near distance), and/or lower (better) stereoscopic acuity thresholds were more accurate on the placement task. No screening measures were predictive of subjective discomfort, perhaps due to the low levels of discomfort induced.					
15. SUBJECT TERMS S3D, stereopsis, optometry, binocular vision, virtual environment, depth perception					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT:	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON (Monitor) Dr. Darrel G. Hopper 19b. TELEPHONE NUMBER (Include Area Code)
a. REPORT	b. ABSTRACT	c. THIS PAGE			
Unclassified	Unclassified	Unclassified	SAR	29	

Technical Paper (TP)

Optometric Measurements Predict Performance but not Comfort on a Virtual Object Placement Task with a Stereoscopic 3D Display

Reprint of paper submitted to: *Optical Engineering* 2014; 53(6):061711.

Authors:

John P. McIntire, PhD, 711 HPW/RHCV, Air Force Research Laboratory
Steven T. Wright, O.D., USAFSAM, Air Force Research Laboratory
Lawrence K. Harrington, O.D., PhD, Ball Aerospace & Technologies Corp.
Paul R. Havig, PhD, 711 HPW/RHCV, Air Force Research Laboratory
Scott N. J. Watamaniuk, PhD, Psychology Department, Wright State University
Eric L. Heft, RHCV, 711 HPW/RHCV, Air Force Research Laboratory

Short Title: Predictors of Performance and Comfort for Stereo 3D Virtual Object Placement

Abstract

Twelve participants were tested on a simple virtual object precision placement task while viewing a stereoscopic 3D (S3D) display. Inclusion criteria included uncorrected or best corrected vision of 20/20 or better in each eye and stereopsis of at least 40 arc sec using the Titmus stereo test. Additionally, binocular function was assessed, including measurements of distant and near phoria (horizontal and vertical) and distant and near horizontal fusion ranges using standard optometric clinical techniques. Before each of six 30 minute experimental sessions, measurements of phoria and fusion ranges were repeated using a Keystone View Telebinocular and an S3D display, respectively. All participants completed experimental sessions in which the task required the precision placement of a virtual object in depth at the same location as a target object. Subjective discomfort was assessed using the Simulator Sickness Questionnaire (SSQ). Individual placement accuracy in S3D trials was significantly correlated with several of the binocular screening outcomes: viewers with larger convergent fusion ranges (measured at near distance), larger total fusion ranges (convergent plus divergent ranges, measured at near distance), and/or lower (better) stereoscopic acuity thresholds were more accurate on the placement task. No screening measures were predictive of subjective discomfort, perhaps due to the low levels of discomfort induced.

Key Words: S3D, stereopsis, optometry, binocular vision, virtual environment, depth perception

Introduction and Background

Stereoscopic 3D (S3D) displays are currently finding wide interest and utility across a variety of task domains, including entertainment, medical, engineering, and military applications. Recently, several of the authors reviewed the state-of-the-art on research and the potential performance benefits of S3D (McIntire, Havig, and Geiselman, 2012, 2014). We found that S3D can improve performance on many different types of spatial tasks involving precision object manipulation (real or virtual), visually finding or identifying objects, navigating, and understanding complex objects or scenes. Further, the benefits provided by S3D are especially apparent for novices in a particular task domain, for difficult or complex spatial tasks, or when other (monocular) cues to depth are degraded or absent. As the popularity and utility of S3D displays continues to expand, and as S3D displays find use in new largely-untested domains, there is growing interest in identifying individuals for whom S3D displays may be of particular value, both in terms of improving performance and ensuring viewing comfort. This could be especially helpful for defining operator selection criteria in occupational fields that rely heavily on S3D viewing, such as robotic surgery, imagery analysis, or aerial refueling, to name but a few.

Predictors of Performance and Comfort on S3D Displays

A variety of research has associated measures of binocular function with S3D viewing, with the goal of identifying objective indicators of visual fatigue or discomfort (e.g., Fortuin, Lambooij, IJsselsteijn, Heynderickx, Edgar, and Evans, 2010; Neveu, Priot, Plantier, and Roumes, 2010). Only a few studies appear to have used these findings to predict individual discomfort on S3D displays (these will be discussed below). To our knowledge, no research has explicitly studied them as possible predictors of individual spatial task performance on S3D displays.

Stereoacuity. We can reasonably suspect *stereoacuity* measures to be predictive of depth task performance. However, the existing literature appears to only have used stereoacuity measures to exclude participants with abnormal or deficient binocular vision or to classify viewers into “good” versus “poor” stereovision groups. Thus, the relationship between individual stereoacuity and performance, particularly for viewers with normal binocular vision, has been largely ignored. In a review of performance issues and

the design of experiments testing stereoscopic 3D displays, Hsu, Pizlo, Chelberg, Babbs, and Delp (1996) had recommended the consideration of individual differences in stereoacuity, and speculated that “depending on the stereo perception task that is required of the subjects, stereoacuity tests may or may not be a good predictor of task performance” (p. 814).

Related previous research on stereoacuity in regards to S3D is sparse and somewhat conflicted. For example, Hale and Stanney (2006) tested two groups in a S3D virtual environment on locomotion, object manipulation, and reaction time tasks. One group had “low” stereo acuity (worse than 80 arc sec) and the other group had “good” stereoacuity (80 arc sec or better). The only notable performance difference between the two groups was that the “good stereoacuity” group made more efficient movements during object manipulation. The primary performance measures were comparable between groups, and the groups’ ratings of post-session discomfort were not significantly different. Other research by Häkkinen, Pölönen, Takatalo, and Nymen (2006) examined sickness/discomfort ratings in S3D using a virtual environment car racing game, but found that individual measurements of stereoacuity were not predictive of individual sickness ratings. Performance scores were apparently not assessed in this research, nor correlated with the individual stereoacuity measures.

Apart from performance specifically on S3D displays, a variety of experiments confirm that stereoacuity plays a key role in performance on real-world depth tasks. For instance, O’Connor et al. (2010) showed that viewers with normal stereoacuity (60 to 250 arc sec or better, depending on the clinical test) generally performed better on pegboard, bead, and water-pouring tasks than those with reduced stereoacuity, and those with reduced stereoacuity often performed better than those with no measurable stereoacuity. Unfortunately, as in most studies, individual stereoacuity was not correlated with individual performance, and viewers with clinically “normal” stereopsis were simply compared (as a group) to non-normal groups.

Fusion Ranges. We might also expect viewers' binocular *fusion ranges* to be related to task performance on S3D displays. Viewers with smaller ranges might have problems fusing larger disparity stimuli, which could manifest as performance deficits on depth-related S3D tasks. Alternatively, viewers

with larger ranges might benefit more generally from the use of S3D displays since they could conceivably fuse a larger range of disparities, without experiencing excessive eyestrain or diplopia.

In the eyestrain and visual fatigue literature, there is some support for a relationship between fusion ranges and viewing discomfort on S3D displays. Nojiri, Yamanoue, Hanazato, Emoto, and Okano (2004) and Emoto, Niida, and Okano (2005) both found that viewers' fusion ranges (i.e., relative vergence limits) decreased after viewing stereoscopic imagery, which presumably indicates that the vergence demands of stereo viewing had adversely fatigued the binocular system. Kim, Choi, Park, and Sohn (2011) demonstrated that viewers with smaller fusion ranges experienced more discomfort from S3D viewing than viewers with larger ranges. Additionally, Neveu, Priot, Plantier, and Roumes (2010) showed that reading text for 10 minutes through a hyperstereoscope (telestereoscope) shifted viewers' binocular fusion limits towards convergence. Research by Chen, Shi, and Tai (2012) and Kim, Choi, and Sohn (2011) demonstrated strong correlations between individuals' binocular fusion limits and their ranges of disparity for comfortable viewing.

Despite these previous studies which suggest a relationship between fusion ranges and comfort, the relationship between fusion limits and S3D performance has received little research attention. One exception is the work by Lambooij, Fortuin, IJsselsteijn, and Heynderickx (2012) who classified participants into two groups based upon a speeded binocular-reading task. The group with "non-normal" reading performance had smaller fusion ranges, reported higher levels of eyestrain caused by S3D display viewing, and had noticeable shifts in their fusion amplitudes from pre- to post-session viewing.

Phorias. Similar to fusion ranges, we might expect viewers' *phorias* to be related to performance and/or comfort on S3D displays, since the vergence eye movements demanded by some S3D stimuli may be far different than a viewer's natural, comfortable resting state of tonic vergence (i.e., sometimes referred to as dissociated phoria or dark vergence). Other researchers have made similar speculations. In regards to eyestrain, Lambooij, Fortuin, IJsselsteijn, Evans, and Heynderickx (2011) hypothesized that it might be particularly difficult for viewers with excessive phorias to accomplish binocular fusion on S3D

displays (their research findings will be described below). It is conceivable that this might also result in performance deficits on S3D display tasks.

There is mixed support for a relationship between phorias and S3D viewing. Häkkinen et al. (2006), as mentioned previously in regards to stereoacuity, also found that while stereo viewing elevated subjective sickness ratings, pre-screening measurements of individuals' horizontal phorias were not predictive of their subsequent sickness ratings. In contrast, it has been demonstrated that viewing stimuli through a hyperstereoscope (telestereoscope) induces changes in viewers' horizontal near and far phorias (Neveu, Priot, Fuchs, and Roumes, 2009; Neveu, Priot, Plantier, and Roumes, 2010). A recent review of eyestrain effects caused by S3D suggested mixed and inconclusive results in regards to whether viewing stereo displays can alter individuals' phorias and thus serve as an objective indicator of eyestrain (Howarth, 2011). A more recent experimental evaluation of this question found that S3D indeed shifted viewer's phorias, but individual phoria changes did not correlate with subjective ratings of discomfort (Karpicka & Howarth, 2013). Given the mixed results in the literature, it remains unclear whether individuals' phorias might serve as a predictor of individual eyestrain or performance on S3D displays.

Other Related Work and the Present Research

Only a few empirical studies have examined several of these clinical screening parameters simultaneously in relation to comfort and/or performance on S3D displays. Lambooij et al. (2011) used a binocular status index classification scheme to group viewers into two groups: those with "good" binocular status, and those with "moderate" binocular status. Their classification scheme took account of viewers' initial dissociated phorias and fusion ranges. The researchers found that the good binocular status group experienced less discomfort and performed better on a stereoscopic reading task relative to 2D reading. But again, individual performance was apparently not studied in relation to individual ophthalmic status. Research by Shibata, Kim, Hoffman, and Banks (2011) measured individuals' dissociated phorias and binocular fusion limits, and found that these predicted individual susceptibility to discomfort on two experiments involving S3D display viewing, at least for near and mid-distance viewing (but not far). No task performance measures were collected in their work.

The present research was an attempt to overcome some of the gaps in the existing literature, and to experimentally examine these three possible clinical predictors (stereoacuity, fusion ranges, and phorias) on both individual performance and comfort. We utilized a simple virtual object precision alignment task conducted on a desktop S3D display.

Methods

Participants. Twelve participants were included in this study, ranging from 19-35 years old, with a mean age of 28 years. The male-to-female ratio was 7:5. All twelve participants had normal or corrected-to-normal distance acuity in both eyes (20/20 or better), and demonstrated at least 40 arc sec of stereoacuity on the Titmus stereovision test. We also measured individuals' refractive errors, binocular fusion ranges (convergence and divergence break and recovery points, at near and far), horizontal phorias (near and far), and vertical phorias (near and far) using standard optometric clinical techniques. Two volunteers were excluded due to reduced stereopsis. A brief demographic and personal history questionnaire related to S3D viewing was also administered, and inter-pupillary distances (IPDs) were measured by the experimenter. All participants read and signed an informed consent document, and the experimental protocol was reviewed, approved, and classified as minimal risk by the Air Force Research Laboratory's Institutional Review Board.

Vision and Comfort Measurements. Prior to each of six experimental sessions, participants' horizontal phorias (near and far) and vertical phoria (far) were measured using the Keystone View Telebinocular vision screening apparatus (Keystone View Company/Mast Development Company; Meadville, Pennsylvania). We also measured participants' fusion ranges for stereoscopic stimuli on the S3D display using a modified method of adjustment: A stimulus at the plane of the display was slowly moved inward using crossed disparity (towards the viewer in depth) until the image either became blurry or broke into a double image, at which point the viewer signaled with a button press. The image was reset at the depth plane of the screen and moved in the opposite direction, again until the image either became blurry or broke into two, and again the viewer signaled this event with a button press. This procedure gave measures of the near and far limits of fusion (convergence and divergence limits, respectively)

which were converted into angular values of binocular disparity (arc deg). Summing these values provided an angular measure of an individual's total fusion range at near distance (the distance of the display).

Visual discomfort (and virtual environment discomfort in general) was measured by administering the standardized subjective questionnaire known as the Simulator Sickness Questionnaire or SSQ (Kennedy, Lane, Berbaum, & Lilienthal, 1993) both before and after each experimental session. Kennedy and colleagues' original analysis of SSQ responses revealed three factors, one of which (the Oculomotor subscale) concerned the ratings on 7 items relating to asthenopia: general discomfort, fatigue, headache, eyestrain, difficulty focusing, difficulty concentrating, and blurred vision. During test administration, each item received a subjective rating score from the participant, ranging from zero (no discomfort) to three (extreme discomfort). These scores were tallied and summed, and the pre-test scores were subtracted from the post-test scores to arrive at difference scores, which indicated changes in discomfort over time in the experimental session. The Oculomotor subscale SSQ difference scores were then averaged across all five S3D sessions per participant, and finally correlated with each participant's optometric measurements of interest. Correlations were tested for significance using *t*-tests.

Display and Apparatus. A high-resolution temporally-multiplexed 120 Hz stereoscopic 3D display was used to present the imagery to the participants (NVIDIA Personal GeForce 3D Vision Active Shutter Glasses, and Samsung® SyncMaster™ 2233RZ). This display was a 22-inch diagonal LCD display with a refresh rate of 120 Hz with native resolution of 1680 (horizontal) x 1050 (vertical). This display system required the wearing of electro-optical active shutter glasses that rapidly oscillated between translucence and opacity in synchrony with the display's oscillation between each eye's imagery (at 60 Hz per eye). For the purpose of this study, observers viewed this display at a distance of approximately 24 inches. A standard QWERTY keyboard and mouse were utilized for the participants' interactions with the display system.

Software. A Microsoft Excel workbook was created to track each participant's progress through the randomized ordering of pre and post-tests. The primary task was written in Visual C++ using the

OpenSceneGraph library to handle creation and manipulation of the viewing volume on the stereoscopic display. Care was taken to match the screen size and viewing distance to the virtual camera and viewing volume. The disparity calculations were verified by placing the test object at a series of distances into and out of the screen, and a high resolution camera captured the resulting left-right image pairs, allowing for on-screen disparities (stereopair half-image separations) to be measured.

Task. The task required the precision placement (spatial alignment) of a virtual object. For each trial, the participant used their right hand to control a computer mouse to position a virtual “control” object (e.g., a small textured pyramid or peg) at an indicated depth on the display, matching the depth and vertical positioning of an identical reference or “target” object. This task served as a replication-and-extension of previous work by Rosenberg (1993) who tested a similar virtual object positioning task and measured alignment accuracy. On each trial, the target object appeared at a randomly chosen point on the target plane. The control object started every trial at the intersection of the control plane and the screen plane, centered along the x-axis. Movement of the control object was limited to the horizontal (x-z) control plane. The target object remained stationary at all times during each trial.

The following magnitudes of the computer-generated stimuli are reported in virtual inches, as the computer model of the task was designed to correspond as accurately as possible with the real-world viewer/display space. The target and control planes were vertically separated by a gap of 2 inches, and measured 8 inches wide by 14 inches deep. The two planes both extended in the z-dimension of virtual space 5.1 inches coming out of the screen, towards the viewer, and 8.8 inches behind the screen away from the viewer. Both the target and control objects were 1 inch tall and 0.56 inches at their widest, and centered vertically in their respective planes, so the vertical separation between the bottom of the control object and the top of the target object was 1 inch. See Figure 1.

Participants pressed the keyboard space bar with their left hand when satisfied with the alignment. Performance measures included completion times and positional error (difference between optimal placement and actual placement in x-z space). Accuracy was emphasized as the primary measure of interest.

Stimuli and Disparities. Binocular disparity limits were fixed within each session to limit the amount of disparity (crossed or uncrossed) on any given trial to a maximum of 0, 20, 40, 60, 80, or 100 arc min. This manipulation was analogous to fixing virtual camera separation in each session to a single value, which differed across sessions. Another analogous way to think about this manipulation is that the virtual IPD ranged from 0 to 100% (assuming an average IPD of 2.6 inches, or 66 mm) in 20% steps of “microstereopsis”, with 0% corresponding to a session with no stereopsis cues, and 100% corresponding to sessions with orthostereoscopic disparity cues. See the Table 1 for comparisons between these equivalent formulations. Each experimental session presented only one limit/range per session. The order in which disparity limits were presented (one per session) was randomized across participants via a Latin Square design.

Procedure. After the brief pre-testing measurements, the 30-minute experimental session began. Trials were entirely self-paced. A total of six sessions (corresponding to the six disparity limit manipulations) were completed by each participant. Each experimental session was completed on a different day. Five-minutes of practice/training were permitted before the start of the first session. Participants on average completed 300 trials per session with an average response time of 6 seconds per trial.

In our analysis for this paper, we excluded performance in the zero-disparity (non-stereo) sessions, as these results will be presented elsewhere as an individualized comparison between non-stereo and S3D (e.g., McIntire, Havig, Harrington, Wright, Watamaniuk, and Heft, 2013). We instead focused the present analysis on performance in only the S3D conditions, because we wished to explore the relationship between clinical screening tests and S3D performance. Each of the participant’s performance was averaged across all their own trials in the five S3D sessions to give each individual an overall measure of S3D placement accuracy (average placement error magnitudes, in units of virtual inches). These measures were then correlated with the individuals’ clinical measures. When a theoretical direction of effect on performance was specifiable a priori (such as the idea that individuals with larger fusion

ranges or smaller refractive errors would perform better on S3D displays), a one-tailed *t*-test was used to test the significance of correlations at the .05 level; otherwise, two-tailed tests were utilized.

Results and Discussion

Predictors of S3D Performance

Table 2 reports the pre-experiment optometric screening tests performed and their correlations with S3D performance. *Near fusion range* was the only finding to demonstrate a significant correlation ($r=-.51$, $p=.045$), as shown graphically in Figure 2. The fusion range is an average of the break and recovery points for both base-in prism and base-out prism, added together to derive a functional range for fusion in units of prism diopter. Participants with a larger fusion range measured at near distance tended to have smaller errors in the S3D placement task. Conceptually, this finding suggests that viewers with larger fusion ranges for near-focused stimuli were able to properly fuse the larger disparities that might be uncomfortable (or impossible) for others to view when using a desktop stereo system at near distance. However, this single result must be considered with caution, because if a family-wise error rate of .05 were applied to this set of 16 tests, none would have achieved a critical *p*-value of $.05/16 = .0031$ or less (although support for this finding is corroborated by similar but even stronger findings in the pre-session measurements, as will be discussed).

It is worth noting that our pre-experimental clinical measure of “fusion range” is not standard, as no standard seems to exist. For each viewer, and for near and far distances, we calculated the fusion range by taking the averaged break and recovery points for convergence and then for divergence, and summing these two values for a total angular extent of fusion. We thought this method advantageously captures, in a single estimate of fusion range ability, four distinct clinical measurements (base-in break, base-in recovery, base-out break, and base-out recovery points). But other calculations utilizing these measurements in a different manner are available in the literature, and in fairness should be compared to our method. For instance, the distance to a single blur-point or breakpoint (convergent or divergent) is often referred to as positive or negative “fusional reserves” (e.g., Endrikhovski, Jin, Miller, & Ford, 2005), “horizontal fusional reserves” (e.g., Fortuin et al., 2010), or “vergence amplitude” (e.g., AOA,

2011). Other authors have used the term “prism vergence amplitude” to describe the total distance between the convergent and divergent breakpoints or, alternatively, the total distance between the convergent and divergent recovery points (e.g., Evans, Drasdo, & Richards, 1994).

For comparison, we provide two of these more common methods for calculating “fusion range” and apply them to the near fusion range-related measures we collected in the clinic. We correlated these two methods with S3D task performance: (1) the total distance between breakpoints, and (2) the total distance between recovery points. We utilized one-tailed *t*-tests for determining significance, as a suspected direction of effect was specifiable a priori. The total distance between breakpoints was not significant ($r=-.22$, $p=.246$) but the distance between recovery points was significantly related to performance ($r=-.71$, $p=.005$). These results suggest that it is the limits of the recovery of fusion (and not the limits of breaking fusion) that were underlying our observation of a relationship between fusion ranges and S3D performance. More speculatively, this might suggest that measures of binocular vision *recovery ranges*, as opposed to the typical *blurring and/or breaking* ranges, may be better able to characterize viewers’ capabilities with S3D stimuli. Conceptually, recovery measures may help identify individuals who can more easily bring diplopic, non-fused stimuli into alignment for fusion to occur as intended, especially large disparity stimuli, if and when such stimuli may appear on an S3D system.

For the five pre-session screening tests (and the one derived measure, fusion range), two were significantly correlated with performance in the suspected direction; see Table 3. These were *fusion near point* ($r=-.50$, $p=.049$) and *fusion range* ($r=-.60$, $p=.020$), shown in Figures 3 and 4, respectively. The pre-session findings were consistent with the pre-experimental screening results shown in Table 2 and described in the preceding paragraphs. We found viewers with closer near points of convergence for S3D stimuli, and viewers with larger fusion ranges, performed better on S3D displays. Fusion ranges, encompassing both near and far fusion breakpoint limits, are plotted by individual participant in rank order according to S3D performance in Figure 5. Again, these relationships suggest that some participants were better able to view the larger disparities on S3D displays without losing fusion, particularly for near

stimuli requiring large convergent eye movements (and inducing larger accommodation-convergence conflicts). Pre-session measures of phorias were not significantly related to performance.

Although all participants scored near-perfect on the pre-screening stereoacuity test (Titmus), indicating stereoacuities of roughly 40 arc sec or better, we also measured stereoacuity thresholds for all participants via follow-up testing. For estimating a true threshold, the Titmus stereotest is typically inadequate because its lower limit is 40 arc sec, but stereoacuities as low as 2-3 arc sec are observable from normal viewers under ideal conditions (Fielder and Moseley, 1996). To obtain stereoacuity estimates for each participant, we used custom adaptive thresholding software that utilizes the QUEST method (Watson and Pelli, 1983). The measurements consisted of 40 trials of near/far forced-choice judgments of a single vertical bar flanked by two reference bars (size and position cues were controlled so that only disparity cues could be used to perform the task). Two of our participants had difficulty with the threshold measurements as conducted on the S3D display: their thresholds indicated stereoacuities many orders of magnitude worse than that indicated by their Titmus tests, and they reported that the central bar was often not perceived in depth even with large disparity magnitudes, or that its perceived position of near versus far alternated over time (indicating an unstable percept). Instead, we were able to easily estimate their stereoacuity thresholds using the Optec Vision Tester (OVT) and/or Randot clinical stereotests (which both test lower values than the Titmus). Neither participant scored 100% on the OVT or Randot tests, indicating that we were reaching their thresholds with these tests (in the 25 to 30 arc sec range for both, which in general are very good values for stereoacuity, but are not excellent).

Observed stereoacuities in the group ranged from 6 to about 30 arc sec (with a mean of 14 arc sec). We correlated the twelve stereoacuity measures with each individual's placement accuracy across all of the S3D trials and found a strong significant correlation ($r=.76$, $p=.002$), shown in Figure 6. Excluding the two participants whose thresholds were estimated using different methods than the rest of the group, we still found a significant correlation ($r=.61$, $p=.031$), despite the drop in sample size. It is also worth noting that these two participants [5 and 7] were two of the three worst performers on the S3D task. If this relationship between stereoacuity and S3D performance holds with larger samples and across different

task types, it may provide a relatively easy-to-administer optometric measure that is predictive of individual task performance on stereoscopic 3D displays.

Predictors of S3D Viewing Comfort

In an attempt to potentially predict which viewers might find S3D displays particularly uncomfortable, we correlated each individual's SSQ self-reported average changes in discomfort (pre-to-post) in the S3D display conditions, with each individual's pre-experiment and pre-session measurements of binocular status. We found no statistically significant correlations between these clinical findings and reported discomfort as induced by the stereo display. This null finding may be due to the fact that relatively low levels of discomfort appear to have been induced in our experiment, perhaps because we utilized short viewing durations (30 minutes) and/or limited the maximum binocular disparity presented in any given trial within a session (at most 100 arc minutes). In the existing literature, both viewing time and the magnitude of binocular disparity are key factors that are often found to effect viewing comfort with S3D displays (e.g., Lambooij, IJsselsteijn, Fortuin, & Heynderickx, 2009; Wöpping, 1995; & Howarth, 2011).

Conclusions and Future Work

In conclusion, we have shown that measures of *total fusion range* (convergence plus divergence limits, measured at near distance), *fusion convergence limits* (measured at near), and *stereoacuity* were useful predictors of placement accuracy performance on a desktop stereoscopic 3D display system. Specifically, viewers with larger total fusion ranges, closer near points of convergence with S3D stimuli, and smaller (better) stereoacuity thresholds generally performed better in terms of accuracy. Total fusion ranges, in particular, were consistently related to performance: this was true when measuring fusion ranges at near distance by standard technique in the clinical setting (utilizing a phoropter and prisms) *and* when repeating related fusion range measurements on the desktop S3D display system before each experimental session.

Our research may be the first to report that for viewers with clinically normal stereopsis, there is a strong significant relationship between stereoacuity and performance on an S3D virtual object precision

placement task. This may also be the first experimental study confirming a relationship between individual fusion limits/ranges and subsequent individual performance on S3D displays. Our results also tentatively suggest the intriguing possibility that it is the *recovery points* of the fusion range, and not necessarily the *breakpoints*, that were driving the correlation between S3D performance and fusion range (as measured clinically). We failed to find any significant correlations between phorias and performance. Future research on the relationship between optometric predictors and S3D performance is recommended to verify these findings.

None of the screening tests were significantly related to the inducement of discomfort on S3D displays, as measured by the pre- to post-session changes in simulator sickness (SSQ) ratings. In the present study, large magnitudes of discomfort were simply not induced by S3D; most viewers found the disparity ranges tested (up to 100 arc min) generally comfortable and usable. Further research on this topic using larger disparity limits, different viewing durations, alternative visual fatigue, eyestrain, and discomfort measures, or other possible optometric predictors may be warranted.

ACKNOWLEDGMENTS

We wish to thank the anonymous reviewers for their helpful comments. Particularly, we would like to acknowledge one reviewer for guidance and clarity provided on the variety of operational definitions of “fusion ranges” that exist in the literature, and for the suggestion of comparing these to our statistical findings.

REFERENCES

AOA (2011). Optometric Clinical Practice Guideline: Care of the Patient with Accommodative and Vergence Dysfunction. Prepared by the American Optometric Association (AOA) Consensus Panel on Care of the Patient with Accommodative and Vergence Dysfunction. Available: <http://www.aoa.org/documents/optometrists/CPG-18.pdf>

- Chen, Z., Shi, J., and Tai, Y. (2012). "An experimental study on the relationship between maximum disparity and comfort disparity in stereoscopic video." *Proceedings of SPIE, 8556, Holography, Diffractive Optics, and Applications V*, 855608.
- Endrikhovski, S., Jin, E., Miller, M. E., & Ford, R. W. (2005). "Predicting individual fusion range from optometric data." *Proceedings of SPIE-IS&T, Electronic Imaging, SPIE 5664*, pp. 36-47.
- Evans, B. J. W., Drasdo, N., & Richards, I. L. (1994). "Investigation of accommodative and binocular function in dyslexia." *Ophthalmic and Physiological Optics*, 14(1), pp. 5-19.
- Fielder, A. R., and Moseley, M. J. (1996). "Does stereopsis matter in humans?" *Eye*, 10, pp. 233-38.
- Fortuin, M. F., Lambooij, M. T., IJsselsteijn, W. A., Heynderickx, I., Edgar, D. F., and Evans, B. J.W. (2010). "An exploration of the initial effects of stereoscopic displays on optometric parameters." *Ophthalmic and Physiological Optics*, 31, pp. 33-44.
- Häkkinen, J., Pölönen, M., Takatalo, J., and Nymen, G. (2006). "Simulator sickness in virtual display gaming: A comparison of stereoscopic and non-stereoscopic situations." *Proceedings of the 8th Conference on Human-Computer Interaction with Mobile Devices and Services*. New York: ACM, pp. 227-230.
- Hale, K. S., and Stanney, K. M. (2006). "Effects of low stereoacuity on performance, presence and sickness within a virtual environment." *Applied Ergonomics*, 37, pp. 329-339.
- Howarth, P. A. (2011). "Potential hazards of viewing 3-D stereoscopic television, cinema and computer games: A review." *Ophthalmic and Physiological Optics*, 31, pp. 111-122.
- Hsu, J., Pizlo, Z., Chelberg, D. M., Babbs, C. F., and Delp, E. J. (1996). "Issues in the design of studies to test the effectiveness of stereo imaging." *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, 26(6), pp. 810-819.
- Karpicka, E., and Howarth, P. A. (2013). "Heterophoria adaptation during the viewing of stereoscopic stimuli." *Ophthalmic & Physiological Optics*, 33(5), pp. 604-610.

- Kennedy, R. S., Lane, N. E., Berbaum, K. S., and Lilienthal, M. G. (1993). "Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness." *International Journal of Aviation Psychology*, 3(3), pp. 203-220.
- Kim, D., Choi, S., Park, S., & Sohn, K. (2011). "Stereoscopic visual fatigue measurement based on fusional response curve and eye-blinks." Proceedings of the 17th International Conference on Digital Signal Processing (DSP), 6-8 July, Corfu, pp. 1-6.
- Kim, D., Choi, S., and Sohn, K. (2011). "Depth adjustment for stereoscopic images and subjective preference evaluation." *Journal of Electronic Imaging*, 20(3), 033011.
- Lambooij, M., Fortuin, M. F., IJsselsteijn, W. A., Evans, B. J. W., and Heynderickx, I. (2011). "Susceptibility to visual discomfort of 3-D displays by visual performance measures." *IEEE Transactions on Circuits and Systems for Video Technology*, 21(12), pp. 1913-1923.
- Lambooij, M., Fortuin, M. F., IJsselsteijn, W. A., and Heynderickx, I. (2012). "Reading performance as a screening tool for visual complaints from stereoscopic content." *Displays*, 33, pp. 84-90.
- Lambooij, M., IJsselsteijn, W. A., Fortuin, M. F., Heynderickx, I. (2009). "Visual discomfort and visual fatigue of stereoscopic displays: A review." *Journal of Image Science and Technology*, 53(3), pp. 1-14.
- McIntire, J. P., Havig, P. R., and Geiselman, E. E. (2012). "What is 3D good for? A review of human performance on stereoscopic 3D displays." Proceedings of SPIE, 8383, 83830X.
- McIntire, J. P., Havig, P. R., and Geiselman, E. E. (2014). "Stereoscopic 3D displays and human performance: A comprehensive review." *Displays*, 35, pp. 18-26.
- McIntire, J., Havig, P., Harrington, L. K., Wright, S., Watamaniuk, S. N. J., and Heft, E. (2013). "Individual differences in performance on a virtual object precision placement task with a stereoscopic 3D display." Unpublished manuscript.
- Neveu, P., Priot, A.-E., Fuchs, P., and Roumes, C. (2009). "Do dynamics models of accommodation/vergence relationship apply for hyperstereoscopic stimuli?" *Perception*, 38, ECVF Abstract Supplement, p. 49.

- Neveu, P., Priot, A.-E., Plantier, J., and Roumes, C. (2010). "Short exposure to telestereo affects the oculomotor system." *Ophthalmic and Physiological Optics*, 30, pp. 806-815.
- O'Connor, A. R., Birch, E. E., Anderson, S., Draper, H., and the FSOS Research Group (2010). "The functional significance of stereopsis." *Investigative Ophthalmology and Visual Science*, 51(4), pp. 2019-2023.
- Rosenberg, L. (1993). "The effect of interocular distance upon operator performance using stereoscopic displays to perform virtual depth tasks." Proceedings of the IEEE VRAIS '93, pp. 27-31.
- Shibata, T., Kim, J., Hoffman, D. M., and Banks, M. S. (2011). "The zone of comfort: Predicting visual discomfort with stereo displays." *Journal of Vision*, 11(8), pp. 1-29.
- Watson, A. B., and Pelli, D. G. (1983). "QUEST: A Bayesian adaptive psychometric method." *Perception and Psychophysics*, 33(2), pp. 113-120.
- Wöpking, M. (1995). "Viewing comfort with stereoscopic pictures: An experimental study on the subjective effects of disparity magnitude and depth of focus." *Journal of the Society for Information Display*, 3, pp. 101-103.

Biographies:

John P. McIntire is an Engineering Research Psychologist at the U.S. Air Force Research Laboratory 711th Human Performance Wing at Wright-Patterson Air Force Base, Ohio. His current research interests include the human factors of 3D displays, network visualizations, and virtual environments. Mr. McIntire holds a Master's degree in Psychology (Human Factors) from Wright State University and is currently pursuing a PhD in Psychology (Human Factors) from Wright State University.

Steve T. Wright received his O.D. degree from Pacific University in 1995 and has worked as clinical and research optometrist with the U.S. Air Force School of Aerospace Medicine since 2004. His research efforts include visual performance after refractive surgery and the occupational assessment of stereopsis and color vision.

Lawrence K. Harrington is a Vision Scientist employed by Ball Aerospace & Technologies Corp. to support battlespace visualization research conducted by the Air Force Research Laboratory. Dr. Harrington is a retired USAF Colonel and holds Doctor of Optometry and Doctor of Philosophy (Physiological Optics) from the University of Missouri-St Louis.

Paul R. Havig is technical advisor for the Battlespace Visualization Branch of the Air Force Research Laboratory. His BA is in Psychology from the University of California at San Diego in 1989 and his MS and PhD in Psychology from the University of Texas at Arlington in 1995 and 1997, respectively. His current research is on the usability of 3D displays, interaction with data in multiple dimensions, and new ways to visualize large data sets.

Scott N.J. Watamaniuk is a professor and Graduate Program Director in the Department of Psychology at Wright State University. He received his BSc in Psychology from the University of Alberta, Edmonton, Alberta, Canada, in 1985 and his MS and PhD in Psychology from Northwestern University in 1987 and 1990, respectively. Dr. Watamaniuk studies the basic processes of human motion perception, pursuit eye movements and their interaction with attention and audition.

Eric L. Heft is the senior software engineer for the Battlespace Visualization Branch of the Air Force Research Laboratory. He received a BS in Computer Engineering from Wright State University in 1995. His current research interests are computer automation, computer vision, identifying and displaying Air Force relevant data on 3D displays.

Table 1. Equivalent formulations of the disparity limits used in the experiment (one limit per session). The binocular disparity limit manipulation can also be considered as manipulation in virtual camera separations, either in raw distance units (mm) or in terms of percentage of a virtual IPD (percentage).

<u>Stereopsis Cues :</u>	<u>none</u>	<u>micro-stereopsis</u>				<u>ortho</u>
Binocular Disparity Limit (arc min)	0	± 20	± 40	± 60	± 80	± 100
Virtual camera separation (vIPD%)	0	20	40	60	80	100
Virtual camera separation (mm)	0.0	13.2	26.4	39.6	52.8	66.0

Table 2. Correlations between the pre-experiment screening tests and S3D performance. One-tailed *t*-tests were performed on the correlations involving refractive errors and fusion ranges with a sample size of 12, while two-tailed tests were used for phorias. Correlations significant at the .05 level are highlighted in grey.

Clinical Measurements	Correlation (<i>r</i>-value)	Significance (<i>p</i>-value)
Refractive Error (right eye)	.31	.163
Refractive Error (left eye)	.13	.344
Horizontal Phoria (distance)	-.29	.361
Vertical Phoria (distance)	-.24	.452
Fusion Range (distance) – Base-In Break	-.16	.310
Fusion Range (distance) – Base-In Recovery	-.21	.256
Fusion Range (distance) – Base-Out Break	-.31	.163
Fusion Range (distance) – Base-Out Recovery	-.10	.379
Fusion Range (distance)	-.38	.112
Horizontal Phoria (near)	-.19	.554
Vertical Phoria (near)	.09	.781
Fusion Range (near) – Base-In Break	-.14	.332
Fusion Range (near) – Base-In Recovery	-.09	.390
Fusion Range (near) – Base-Out Break	-.09	.390
Fusion Range (near) – Base-Out Recovery	-.40	.099
Fusion Range (near)	-.51	.045

Table 3. Correlations between the Pre-session Repeated Screening Tests and S3D Performance. One-tailed *t*-tests were performed on the correlations involving fusion limits with a sample size of 12, while two-tailed tests were used for phorias. Correlations significant at the .05 level are highlighted in grey.

Pre-session Measurements (repeated before each session)	Correlation (<i>r</i>-value)	Significance (<i>p</i>-value)
Lateral Phoria (near)	-.45	.142
Lateral Phoria (far)	-.50	.098
Vertical Phoria (far)	-.45	.142
Fusion Near Limit (Convergence)	-.50	.049
Fusion Far Limit (Divergence)	.34	.860
Fusion Range (Total; Convergence plus Divergence)	-.60	.020

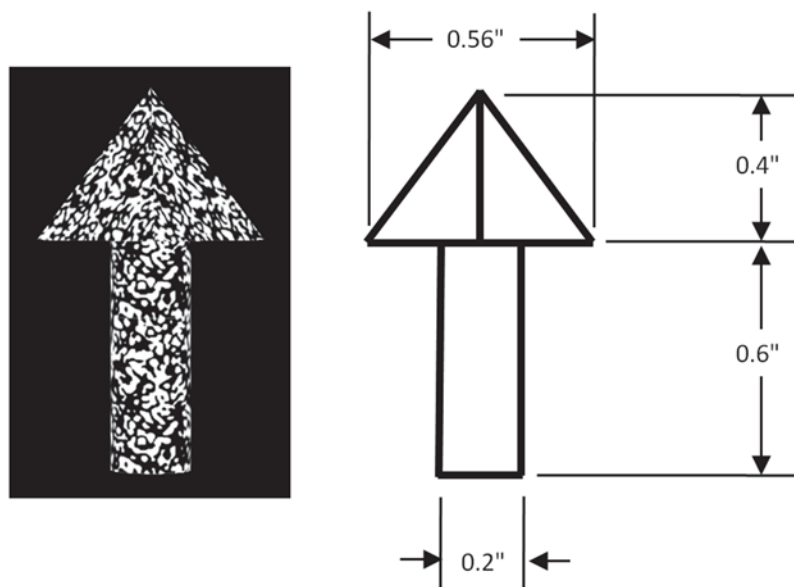
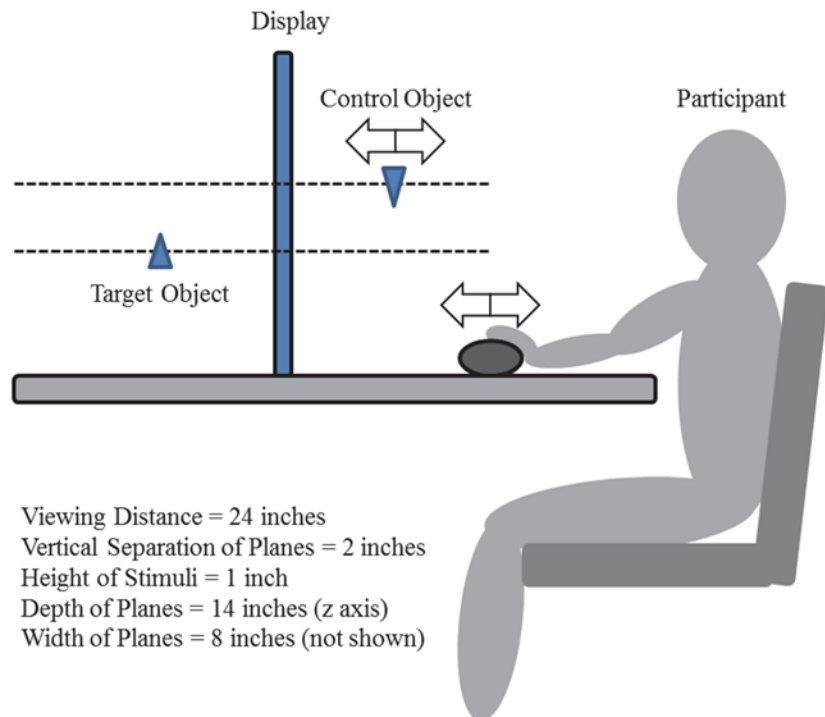


Figure 1. (Top): Schematic side view of the experimental set-up. The participant physically controlled a computer mouse to move the control object within the virtual volume, presented to the viewer via the display. Movement of the control object was limited to the control plane. The task required the precise alignment of the control object overtop the target object. (Bottom): A screen-shot and a schematic side view of the target/control objects, which were small textured arrows or “pegs” consisting of a cylinder with a four-sided pyramid situated on one end.

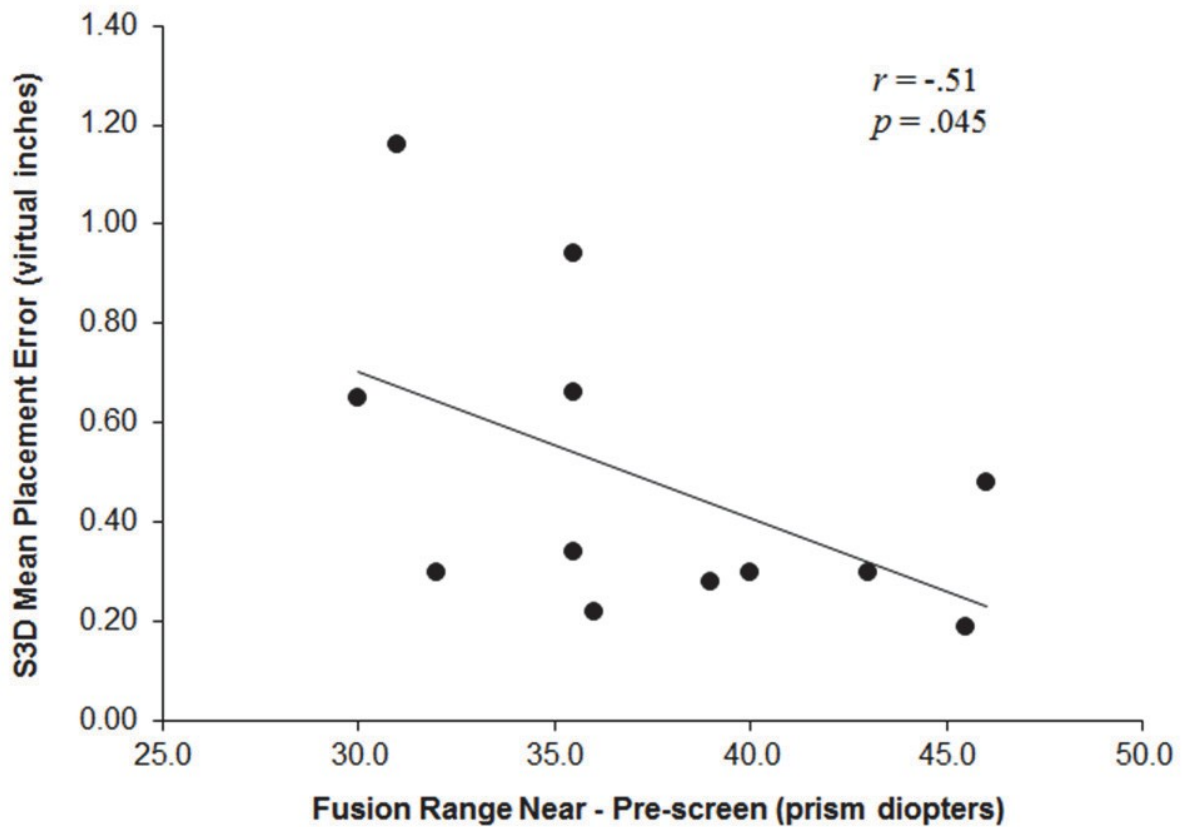


Figure 2. The relationship between the near fusion range (in prism diopters) and placement error performance on the S3D display. Each data point represents the single pre-screening measurement of their fusion range at near for each individual participant.

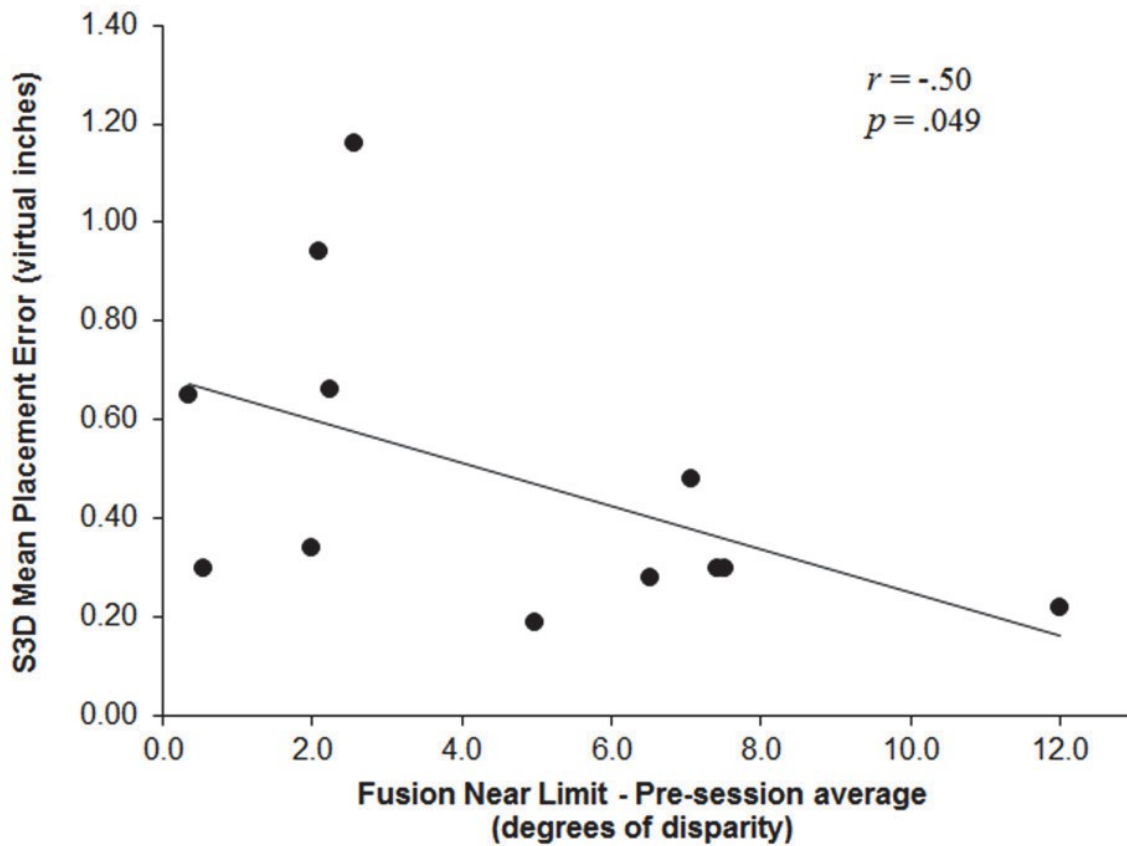


Figure 3. The relationship between the pre-session average measures of fusion near limit (convergence) and placement error performance on the S3D display. Each data point represents each individual's average of six different measurements of their fusion near limits, taken before each experimental session on the S3D display.

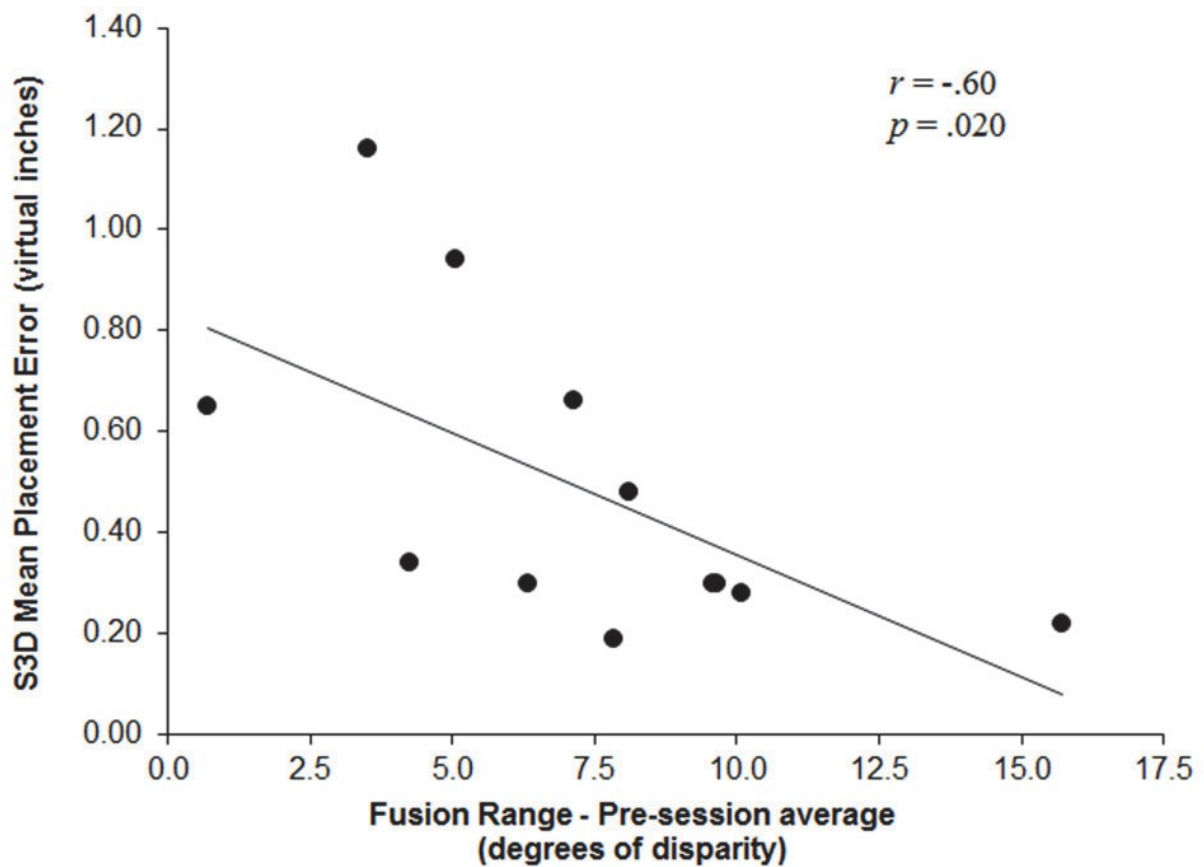


Figure 4. The relationship between the pre-session average measures of fusion range (total; convergence plus divergence limits) and placement error performance on the S3D display. Each data point represents each individual's average of six different measurements of their fusion limits, taken before each experimental session on the S3D display.

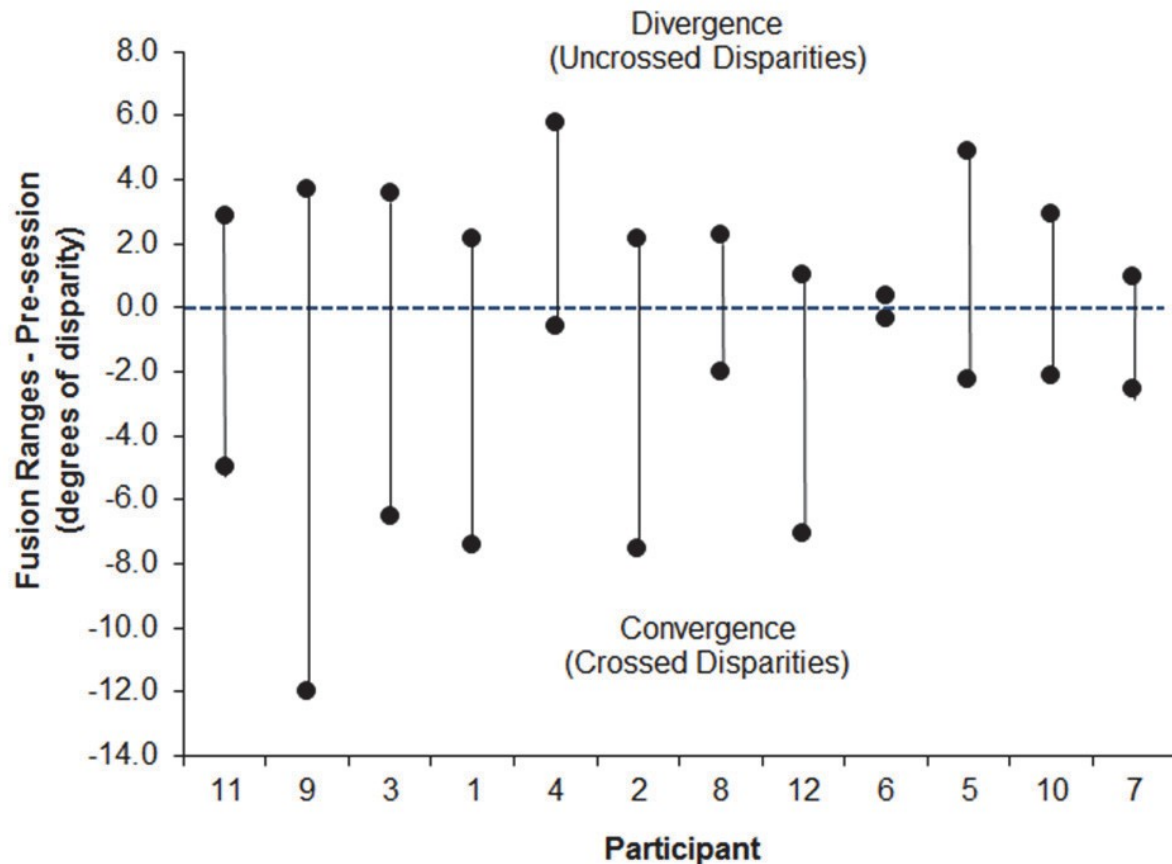


Figure 5. Individuals' pre-session average measures of fusion ranges (total range; convergence to divergence limits), in rank order by placement error performance on the S3D display (best performers from left to right). Each data point represents each individual's average of six different measurements of their fusion limits, taken on the S3D display before each experimental session. Negative fusion values represent crossed screen disparities (convergence limits), and positive fusion values represent uncrossed disparities (divergence limits). The dashed line at zero represents the zero-disparity display surface. Participants with larger total fusion ranges and larger crossed (convergent) fusion ranges generally demonstrated better performance.

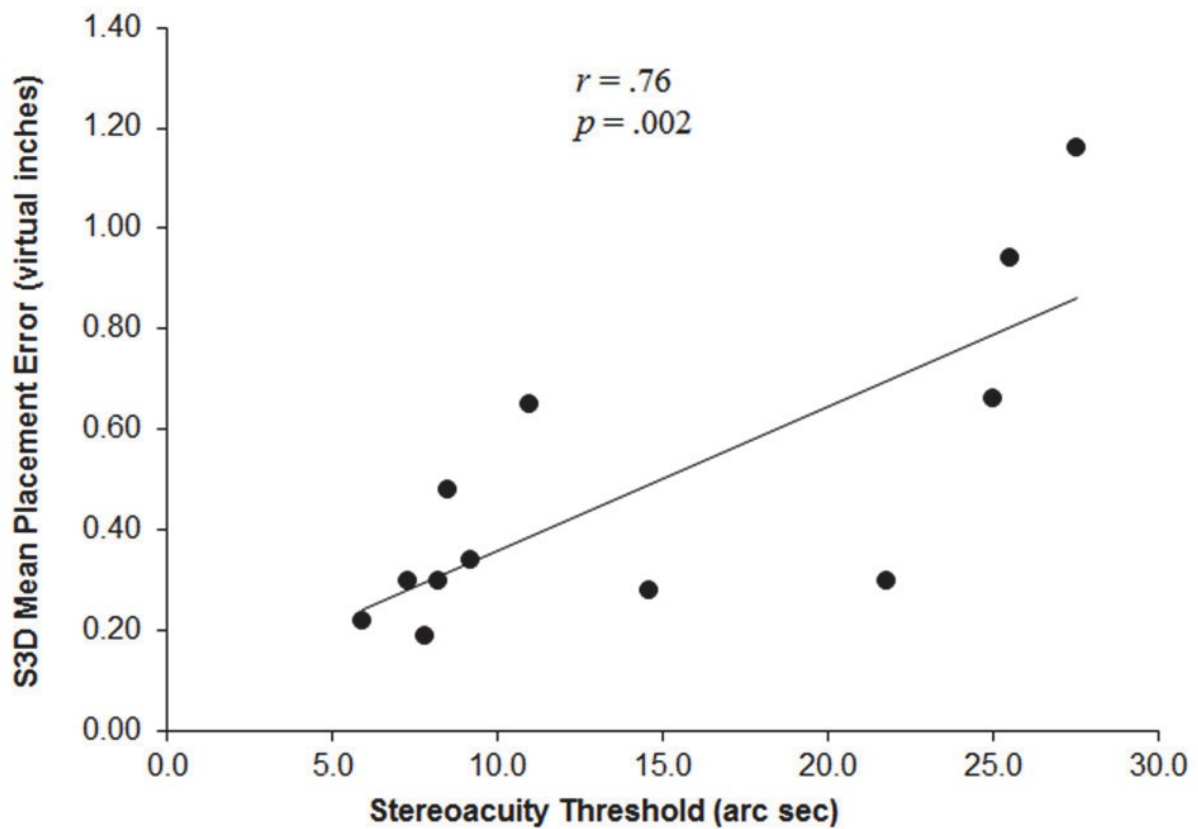


Figure 6. The relationship between stereoacuity thresholds (in arc seconds) and placement error performance on the S3D display. See the text for details on our threshold estimation method and our difficulty in estimating two participants' stereoacuities.